

Communication over Finite-Field Matrix Channels

Danilo Silva, Frank R. Kschischang, and Ralf Kötter

Abstract

This paper is motivated by the problem of error control in network coding when errors are introduced in a random fashion (rather than chosen by an adversary). An additive-multiplicative matrix channel is considered as a model for random network coding. The model assumes that n packets of length m are transmitted over the network, and up to t erroneous packets are randomly chosen and injected into the network. Upper and lower bounds on capacity are obtained for any channel parameters, and asymptotic expressions are provided in the limit of large field or matrix size. A simple coding scheme is presented that achieves capacity in both limiting cases. The scheme has decoding complexity $O(n^2m)$ and a probability of error that decreases exponentially both in the packet length and in the field size in bits. Extensions of these results for coherent network coding are also presented.

Index Terms

Error correction, error trapping, matrix channels, network coding, one-shot codes, probabilistic error model.

I. INTRODUCTION

Linear network coding [1]–[3] is a promising new approach to information dissemination over networks. The fact that packets may be linearly combined at intermediate nodes affords, in many useful scenarios, higher rates than conventional routing approaches. If the linear combinations are chosen in a random,

This work was supported by CAPES Foundation, Brazil, and by the Natural Sciences and Engineering Research Council of Canada. The material in this paper was presented in part at the 24th Biennial Symposium on Communications, Kingston, Canada, June 2008.

D. Silva and F. R. Kschischang are with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: danilo@comm.utoronto.ca, frank@comm.utoronto.ca).

R. Kötter is deceased.

distributed fashion, then random linear network coding [4] not only maintains most of the benefits of linear network coding, but also affords a remarkable simplicity of design that is practically very appealing.

However, linear network coding has the intrinsic drawback of being extremely sensitive to error propagation. Due to packet mixing, a single corrupt packet has the potential to contaminate all packets received by a destination node. The problem is better understood by looking at a matrix model for (single-source) linear network coding, given by

$$Y = AX + DZ. \quad (1)$$

All matrices are over a finite field. Here, X is an $n \times m$ matrix whose rows are packets transmitted by the source node, Y is an $N \times m$ matrix whose rows are the packets received by a (specific) destination node, and Z is a $t \times m$ matrix whose rows are the additive error packets injected at some network links. The matrices A and D are transfer matrices that describe the linear transformations incurred by packets on route to the destination. Such linear transformations are responsible for the (unconventional) phenomenon of error propagation.

There has been an increasing amount of research on error control for network coding, with results naturally depending on the specific channel model used, i.e., the joint statistics of A , D and Z given X . Under a worst-case (or adversarial) error model, the work in [5], [6] (together with [7]–[10]) has obtained the maximum achievable rate for a wide range of conditions. If A is square ($N = n$) and nonsingular, and $m \geq n$, then the maximum information rate that can be achieved in a single use of the channel is exactly $n - 2t$ packets when A is known at the receiver, and approximately $\frac{m-n}{m}(n - 2t)$ packets when A is unknown. These approaches are inherently pessimistic and share many similarities with classical coding theory.

Recently, Montanari and Urbanke [11] brought the problem to the realm of information theory by considering a probabilistic error model. Their model assumes, as above, that A is invertible and $m \geq n$; in addition, they assume that the matrix DZ is chosen uniformly at random among all $n \times m$ matrices of rank t . For such a model and, under the assumption that the transmitted matrix X must contain an $n \times n$ identity submatrix as a header, they compute the maximal mutual information in the limit of large matrix size—approximately $\frac{m-n-t}{m}(n - t)$ packets per channel use. They also present an iterative coding scheme with decoding complexity $O(n^3m)$ that asymptotically achieves this rate.

The present paper is motivated by [11], and by the challenge of computing or approximating the actual channel capacity (i.e., without any prior assumption on the input distribution) for any channel parameters (i.e., not necessarily in the limit of large matrix size). Our contributions can be summarized as follows:

- Assuming that the matrix A is a constant known to the receiver, we compute the exact channel capacity for any channel parameters. We also present a simple coding scheme that asymptotically achieves capacity in the limit of large field or matrix size.
- Assuming that the matrix A is chosen uniformly at random among all nonsingular matrices, we compute upper and lower bounds on the channel capacity for any channel parameters. These bounds are shown to converge asymptotically in the limit of large field or matrix size. We also present a simple coding scheme that asymptotically achieves capacity in both limiting cases. The scheme has decoding complexity $O(n^2m)$ and a probability of error that decays exponentially fast both in the packet length and in the field size in bits.
- We present several extensions of our results for situations where the matrices A , D and Z may be chosen according to more general probability distributions.

A main assumption that underlies this paper (even the extensions mentioned above) is that the transfer matrix A is always invertible. One might question whether this assumption is realistic for actual network coding systems. For instance, if the field size is small, then random network coding may not produce a nonsingular A with high probability. We believe, however, that removing this assumption complicates the analysis without offering much insight. Under an *end-to-end coding* (or *layered*) approach, there is a clear separation between the network coding protocol—which induces a matrix channel—and the error control techniques applied at the source and destination nodes. In this case, it is reasonable to assume that network coding system will be designed to be *feasible* (i.e., able to deliver X to all destinations) when no errors occur in the network. Indeed, a main premise of linear network coding is that the field size is sufficiently large in order to allow a feasible network code. Thus, the results of this paper may be seen as conditional on the network coding layer being successful in its task.

The remainder of this paper is organized as follows. In Section II, we provide general considerations on the type of channels studied in this paper. In Section III, we address a special case of (1) where A is random and $t = 0$, which may be seen as a model for random network coding without errors. In Section IV, we address a special case of (1) where A is the identity matrix. This channel may be seen as a model for network coding with errors when A is known at the receiver, since the receiver can always compute $A^{-1}Y$. The complete channel with a random, unknown A is addressed Section V, where we make crucial use of the results and intuition developed in the previous sections. Section VI discusses possible extensions of our results, and Section VII presents our conclusions.

We will make use of the following notation. Let \mathbb{F}_q be the finite field with q elements. We use $\mathbb{F}_q^{n \times m}$ to denote the set of all $n \times m$ matrices over \mathbb{F}_q and $\mathcal{T}_{n \times m, t}(\mathbb{F}_q)$ to denote the set of all $n \times m$ matrices of

rank t over \mathbb{F}_q . We shall write simply $\mathcal{T}_{n \times m, t} = \mathcal{T}_{n \times m, t}(\mathbb{F}_q)$ when the field \mathbb{F}_q is clear from the context. We also use the notation $\mathcal{T}_{n \times m} = \mathcal{T}_{n \times m, \min\{n, m\}}$ for the set of all full-rank $n \times m$ matrices. The $n \times m$ all-zero matrix and the $n \times n$ identity matrix are denoted by $0_{n \times m}$ and $I_{n \times n}$, respectively, where the subscripts may be omitted when there is no risk of confusion. The reduced row echelon (RRE) form of a matrix M will be denoted by $\text{RRE}(M)$.

II. MATRIX CHANNELS

For clarity and consistency of notation, we recall a few definitions from information theory [12].

A discrete channel $(\mathcal{X}, \mathcal{Y}, p_{Y|X})$ consists of an input alphabet \mathcal{X} , an output alphabet \mathcal{Y} , and a conditional probability distribution $p_{Y|X}$ relating the channel input $X \in \mathcal{X}$ and the channel output $Y \in \mathcal{Y}$. An (M, ℓ) code for a channel $(\mathcal{X}, \mathcal{Y}, p_{Y|X})$ consists of an encoding function $\{1, \dots, M\} \rightarrow \mathcal{X}^\ell$ and a decoding function $\mathcal{Y}^\ell \rightarrow \{1, \dots, M, f\}$, where f denotes a decoding failure. It is understood that an (M, ℓ) code is applied to the ℓ th extension of the discrete memoryless channel $(\mathcal{X}, \mathcal{Y}, p_{Y|X})$. A rate R (in bits) is said to be achievable if there exists a sequence of $(\lceil 2^{\ell R} \rceil, \ell)$ codes such that decoding is unsuccessful (either an error or a failure occurs) with probability arbitrarily small as $\ell \rightarrow \infty$. The capacity of the channel is the supremum of all achievable rates. It is well-known that the capacity is given by

$$C = \max_{p_X} I(X; Y)$$

where p_X denotes the input distribution.

Here, we are interested in matrix channels, i.e., channels for which both the input and output variables are matrices. In particular, we are interested in a family of additive matrix channels given by the channel law

$$Y = AX + DZ \tag{2}$$

where $X, Y \in \mathbb{F}_q^{n \times m}$, $A \in \mathbb{F}_q^{n \times n}$, $D \in \mathbb{F}_q^{n \times t}$, $Z \in \mathbb{F}_q^{t \times m}$, and X , (A, D) and Z are statistically independent. Since the capacity of a matrix channel naturally scales with nm , we also define a *normalized capacity*

$$\overline{C} = \frac{1}{nm} C.$$

In the following, we assume that statistics of A , D and Z are given for all q, n, m, t . In this case, we may denote a matrix channel simply by the tuple (q, n, m, t) , and we may also indicate this dependency in both C and \overline{C} . We now define two limiting forms of a matrix channel (strictly speaking, of a sequence

of matrix channels). The first form, which we call the *infinite-field-size channel*, is obtained by taking $q \rightarrow \infty$. The capacity of this channel is given by

$$\lim_{q \rightarrow \infty} \frac{1}{\log_2 q} C(q, n, m, t)$$

represented in q -ary units per channel use. The second form, which we call the *infinite-rank channel*, is obtained by setting $t = \tau n$ and $n = \lambda m$, and taking $m \rightarrow \infty$. The normalized capacity of this channel is given by

$$\lim_{m \rightarrow \infty} \frac{1}{\log_2 q} \bar{C}(q, \lambda m, m, \tau \lambda m)$$

represented in q -ary units per transmitted q -ary symbol. We will hereafter assume that logarithms are taken to the base q and omit the factor $\frac{1}{\log_2 q}$ from the above expressions.

Note that, to achieve the capacity of an infinite-field-size channel (similarly for an infinite-rank channel), one should find a two-dimensional family of codes: namely, a sequence of codes with increasing block length ℓ for each q , as $q \rightarrow \infty$ (or for each m , as $m \rightarrow \infty$).

We will simplify our task here by considering only codes with block length $\ell = 1$, which we call *one-shot codes*. We will show, however, that these codes can achieve the capacity of both the infinite-field-size and the infinite-rank channels, at least for the classes of channels considered here. In other words, one-shot codes are asymptotically optimal as either $q \rightarrow \infty$ or $m \rightarrow \infty$.

For completeness, we define also two more versions of the channel: the *infinite-packet-length channel*, obtained by fixing q , t and n , and letting $m \rightarrow \infty$, and the *infinite-batch-size channel*, obtained by fixing q , t and m , and letting $n \rightarrow \infty$. These channels are discussed in Section VI-E.

It is important to note that a $(q, n, \ell m, t)$ channel is not the same as the ℓ -extension of a (q, n, m, t) channel. For instance, the 2-extension of a (q, n, m, t) channel has channel law

$$(Y_1, Y_2) = (A_1 X_1 + D_1 Z_1, A_2 X_2 + D_2 Z_2)$$

where $(X_1, X_2) \in (\mathbb{F}_q^{n \times m})^2$, and (A_1, D_1, Z_1) and (A_2, D_2, Z_2) correspond to independent realizations of a (q, n, m, t) channel. This is not the same as the channel law for a $(q, n, 2m, t)$ channel,

$$\begin{bmatrix} Y_1 & Y_2 \end{bmatrix} = A_1 \begin{bmatrix} X_1 & X_2 \end{bmatrix} + D_1 \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$$

since (A_2, D_2) may not be equal to (A_1, D_1) .

To the best of our knowledge, the ℓ -extension of a (q, n, m, t) channel has not been considered in previous works, with the exception of [13]. For instance, [14] and [11] consider only limiting forms of a (q, n, m, t) channel. Although both models are referred to simply as “random linear network coding,”

the model implied by the results in [11] is in fact an infinite-rank channel, while the model implied by the results in [14] is an infinite-packet-length-infinite-field-size channel.

We now proceed to investigating special cases of (2), by considering specific statistics for A , D and Z .

III. THE MULTIPLICATIVE MATRIX CHANNEL

We define the *multiplicative matrix channel* (MMC) by the channel law

$$Y = AX$$

where $A \in \mathcal{T}_{n \times n}$ is chosen uniformly at random among all $n \times n$ nonsingular matrices, and independently from X . Note that the MMC is a $(q, n, m, 0)$ channel.

A. Capacity and Capacity-Achieving Codes

In order to find the capacity of this channel, we will first solve a more general problem.

Proposition 1: Let \mathcal{G} be a finite group that acts on a finite set \mathcal{S} . Consider a channel with input variable $X \in \mathcal{S}$ and output variable $Y \in \mathcal{S}$ given by $Y = AX$, where $A \in \mathcal{G}$ is drawn uniformly at random and independently from X . The capacity of this channel, in bits per channel use, is given by

$$C = \log_2 |\mathcal{S}/\mathcal{G}|$$

where $|\mathcal{S}/\mathcal{G}|$ is the number of equivalence classes of \mathcal{S} under the action of \mathcal{G} . Any complete set of representatives of the equivalence classes is a capacity-achieving code.

Proof: For each $x \in \mathcal{S}$, let $\mathcal{G}(x) = \{gx \mid g \in \mathcal{G}\}$ denote the orbit of x under the action of \mathcal{G} . Recall that $\mathcal{G}(y) = \mathcal{G}(x)$ for all $y \in \mathcal{G}(x)$ and all $x \in \mathcal{S}$, that is, the orbits form equivalence classes.

For $y \in \mathcal{G}(x)$, let $\mathcal{G}_{x,y} = \{g \in \mathcal{G} \mid gx = y\}$. By a few manipulations, it is easy to show that $|\mathcal{G}_{x,y}| = |\mathcal{G}_{x,y'}|$ for all $y, y' \in \mathcal{G}(x)$. Since A has a uniform distribution, it follows that $P[Y = y \mid X = x] = 1/|\mathcal{G}(x)|$, for all $y \in \mathcal{G}(x)$.

For any $x \in \mathcal{S}$, consider the same channel but with the input alphabet restricted to $\mathcal{G}(x)$. Note that the output alphabet will also be restricted to $\mathcal{G}(x)$. This is a $|\mathcal{G}(x)|$ -ary channel with uniform transition probabilities; thus, the capacity of this channel is 0. Now, the overall channel can be considered as a sum (union of alphabets) of all the restricted channels. The capacity of a sum of M channels with capacities C_i , $i = 1, \dots, M$, is known to be $\log_2 \sum_{i=1}^M 2^{C_i}$ bits. Thus, the capacity of the overall channel is $\log_2 M$ bits, where $M = |\mathcal{S}/\mathcal{G}|$ is the number of orbits. A capacity-achieving code (with block length 1) may be obtained by simply selecting one representative from each equivalence class. ■

Proposition 1 shows that in a channel induced by a group action, where the group elements are selected uniformly at random, the receiver cannot distinguish between transmitted elements that belong to the same equivalence class. Thus, the transmitter can only communicate the choice of a particular equivalence class.

Returning to our original problem, we have $\mathcal{S} = \mathbb{F}_q^{n \times m}$ and $\mathcal{G} = \mathcal{T}_{n \times n}$ (the general linear group $GL_n(\mathbb{F}_q)$). The equivalence classes of \mathcal{S} under the action of \mathcal{G} are the sets of matrices that share the same row space. Thus, we can identify each equivalence class with a subspace of \mathbb{F}_q^m of dimension at most n . Let the Gaussian coefficient

$$\begin{bmatrix} m \\ k \end{bmatrix}_q = \prod_{i=0}^{k-1} (q^m - q^i) / (q^k - q^i)$$

denote the number of k -dimensional subspaces of \mathbb{F}_q^m . We have the following corollary of Proposition 1.

Corollary 2: The capacity of the MMC, in q -ary units per channel use, is given by

$$C_{\text{MMC}} = \log_q \sum_{k=0}^n \begin{bmatrix} m \\ k \end{bmatrix}_q.$$

A capacity-achieving code $\mathcal{C} \subseteq \mathbb{F}_q^{n \times m}$ can be obtained by ensuring that each k -dimensional subspace of \mathbb{F}_q^m , $k \leq n$, is the row space of some unique $X \in \mathcal{C}$.

Note that Corollary 2 reinforces the idea introduced in [9] that, in order to communicate under random network coding, the transmitter should encode information in the choice of a subspace.

We now compute the capacity for the two limiting forms of the channel, as discussed in Section II. We have the following result.

Proposition 3: Let $\lambda = n/m$ and assume $0 < \lambda \leq 1/2$. Then

$$\lim_{q \rightarrow \infty} C_{\text{MMC}} = (m - n)n \quad (3)$$

$$\lim_{\substack{m \rightarrow \infty \\ n = \lambda m}} \overline{C_{\text{MMC}}} = 1 - \lambda. \quad (4)$$

Proof: First, observe that

$$\begin{bmatrix} m \\ n^* \end{bmatrix}_q < \sum_{k=0}^n \begin{bmatrix} m \\ k \end{bmatrix}_q < (n+1) \begin{bmatrix} m \\ n^* \end{bmatrix}_q \quad (5)$$

where $n^* = \min\{n, \lfloor m/2 \rfloor\}$. Using the fact that [9]

$$q^{(m-k)k} < \begin{bmatrix} m \\ k \end{bmatrix}_q < 4q^{(m-k)k} \quad (6)$$

it follows that

$$(m - n^*)n^* < C_{\text{MMC}} < (m - n^*)n^* + \log_q 4(n+1). \quad (7)$$

The last term on the right vanishes on both limiting cases. ■

The case $\lambda \geq 1/2$ can also be readily obtained but is less interesting since, in practice, the packet length m will be much larger than the number of packets n .

Note that an expression similar to (7) has been found in [13] under a different assumption on the transfer matrix (namely, that A is uniform on $\mathbb{F}_q^{n \times n}$). It is interesting to note that, also in that case, the same conclusion can be reached about the sufficiency of transmitting subspaces [13].

An intuitive way to interpret (3) is the following: out of the nm symbols obtained by the receiver, n^2 of these symbols are used to describe A , while the remaining ones are used to communicate X .

It is interesting to note that (3) precisely matches (4) after normalizing by the total number of transmitted symbols, nm .

Both limiting capacity expressions (3) and (4) can be achieved using a simple coding scheme where an $n \times (m - n)$ data matrix U is concatenated on the left with an $n \times n$ identity matrix I , yielding a transmitted matrix $X = \begin{bmatrix} I & U \end{bmatrix}$. The first n symbols of each transmitted packet may be interpreted as pilot symbols used to perform “channel sounding”. Note that this is simply the standard way of using random network coding [15].

IV. THE ADDITIVE MATRIX CHANNEL

We define the *additive matrix channel* (AMC) according to

$$Y = X + W$$

where $W \in \mathcal{T}_{n \times m, t}$ is chosen uniformly at random among all $n \times m$ matrices of rank t , independently from X . Note that the AMC is a (q, n, m, t) channel with $D \in \mathcal{T}_{n \times t}$ and $Z \in \mathcal{T}_{t \times m}$ uniformly distributed, and $A = I$.

A. Capacity

The capacity of the AMC is computed in the next proposition.

Proposition 4: The capacity of the AMC is given by

$$C_{\text{AMC}} = nm - \log_q |\mathcal{T}_{n \times m, t}|.$$

For $\lambda = n/m$ and $\tau = t/n$, we have the limiting expressions

$$\lim_{q \rightarrow \infty} C_{\text{AMC}} = (m - t)(n - t) \tag{8}$$

$$\lim_{\substack{m \rightarrow \infty \\ n = \lambda m \\ t = \tau n}} \overline{C_{\text{AMC}}} = (1 - \lambda\tau)(1 - \tau). \tag{9}$$

Proof: To compute the capacity, we expand the mutual information

$$I(X; Y) = H(Y) - H(Y|X) = H(Y) - H(W)$$

where the last equality holds because X and W are independent. Note that $H(Y) \leq nm$, and the maximum is achieved when Y is uniform. Since $H(W)$ does not depend on the input distribution, we can maximize $H(Y)$ by choosing, e.g., a uniform p_X .

The entropy of W is given by $H(W) = \log_q |\mathcal{T}_{n \times m, t}|$. The number of $n \times m$ matrices of rank t is given by [16, p. 455]

$$|\mathcal{T}_{n \times m, t}| = \frac{|\mathcal{T}_{n \times t}| |\mathcal{T}_{t \times m}|}{|\mathcal{T}_{t \times t}|} = |\mathcal{T}_{n \times t}| \begin{bmatrix} m \\ t \end{bmatrix}_q \quad (10)$$

$$= q^{(n+m-t)t} \prod_{i=0}^{t-1} \frac{(1 - q^{i-n})(1 - q^{i-m})}{(1 - q^{i-t})}. \quad (11)$$

Thus,

$$\begin{aligned} C_{\text{AMC}} &= nm - \log_q |\mathcal{T}_{n \times m, t}| \\ &= (m-t)(n-t) + \log_q \prod_{i=0}^{t-1} \frac{(1 - q^{i-n})(1 - q^{i-m})}{(1 - q^{i-t})} \end{aligned}$$

The limiting expressions (8) and (9) follow immediately from the equation above. ■

Remark: The expression (9), which gives the capacity of the infinite-rank AMC, has been previously obtained in [11] for a channel that is equivalent to the AMC. Our proof is a simple extension of the proof in [11].

As can be seen from (11), an $n \times m$ matrix of rank t can be specified with approximately $(n+m-t)t$ symbols. Thus, the capacity (8) can be interpreted as the number of symbols conveyed by Y minus the number of symbols needed to describe W .

Note that, as in Section III, the normalized capacities of the infinite-field-size AMC and the infinite-rank AMC are the same. An intuitive explanation might be the fact that, for the two channels, both the number of bits per row and the number of bits per column tend to infinity. In contrast, the normalized capacity is different when only one of these quantities grows while the other is fixed. This is the case of the infinite-packet-length AMC and the infinite-batch-size AMC, which are studied in Section VI-E.

B. A Coding Scheme

We now present an efficient coding scheme that achieves (8) and (9). The scheme is based on an “error trapping” strategy.

Let $U \in \mathbb{F}_q^{(n-v) \times (m-v)}$ be a data matrix, where $v \geq t$. A codeword X is formed by adding all-zero rows and columns to U so that

$$X = \begin{bmatrix} 0_{v \times v} & 0_{v \times (m-v)} \\ 0_{(n-v) \times v} & U \end{bmatrix}.$$

These all-zero rows and columns may be interpreted as the “error traps.” Clearly, the rate of this scheme is $R = (n - v)(m - v)$.

Since the noise matrix W has rank t , we can write it as

$$W = BZ = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$$

where $B_1 \in \mathbb{F}_q^{v \times t}$, $B_2 \in \mathbb{F}_q^{(n-v) \times t}$, $Z_1 \in \mathbb{F}_q^{t \times v}$ and $Z_2 \in \mathbb{F}_q^{t \times (m-v)}$. The received matrix Y is then given by

$$Y = X + W = \begin{bmatrix} B_1 Z_1 & B_1 Z_2 \\ B_2 Z_1 & U + B_2 Z_2 \end{bmatrix}.$$

We define an error trapping failure to be the event that $\text{rank } B_1 Z_1 < t$. Intuitively, this corresponds to the situation where either the row space or the column space of the error matrix has not been “trapped”.

For now, assume that the error trapping is successful, i.e., $\text{rank } B_1 = \text{rank } Z_1 = t$. Consider the submatrix corresponding to the first v columns of Y . Since $\text{rank } B_1 Z_1 = t$, the rows of $B_2 Z_1$ are completely spanned by the rows of $B_1 Z_1$. Thus, there exists some matrix \bar{T} such that $B_2 Z_1 = \bar{T} B_1 Z_1$. But $(B_2 - \bar{T} B_1) Z_1 = 0$ implies that $B_2 - \bar{T} B_1 = 0$, since Z_1 has full row rank. It follows that

$$T \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \quad \text{where } T = \begin{bmatrix} I & 0 \\ \bar{T} & I \end{bmatrix}.$$

Note also that $TX = X$. Thus,

$$TY = TX + TW = \begin{bmatrix} B_1 Z_1 & B_1 Z_2 \\ 0 & U \end{bmatrix}$$

from which the data matrix U can be readily obtained.

The complexity of the scheme is computed as follows. In order to obtain \bar{T} , it suffices to perform Gaussian elimination on the left $n \times v$ submatrix of Y , for a cost of $O(nv^2)$ operations. The data matrix can be extracted by multiplying \bar{T} with the top right $v \times (n - v)$ submatrix of Y , which can be accomplished in $O((n - v)v(m - v))$ operations. Thus, the overall complexity of the scheme is $O(nmv)$ operations in \mathbb{F}_q .

Note that $B_1 Z_1$ is available at the receiver as the top-left submatrix of Y . Moreover, the rank of $B_1 Z_1$ is already computed during the Gaussian elimination step of the decoding. Thus, the event that the error

trapping fails can be readily detected at the receiver, which can then declare a decoding failure. It follows that the error probability of the scheme is zero.

Let us now compute the probability of decoding failure. Consider, for instance, $P_1 = P[\text{rank } Z_1 = t]$, where $Z = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$ is a full-rank matrix chosen uniformly at random. An equivalent way of generating Z is to first generate the entries of a matrix $M \in \mathbb{F}_q^{t \times m}$ uniformly at random, and then discard M if it is not full-rank. Thus, we want to compute $P_1 = P[\text{rank } M_1 = t \mid \text{rank } M = t]$, where M_1 corresponds to the first v columns of M . This probability is

$$\begin{aligned} P_1 &= \frac{P[\text{rank } M_1 = t]}{P[\text{rank } M = t]} = \frac{q^{mt} \prod_{i=0}^{t-1} (q^v - q^i)}{q^{vt} \prod_{i=0}^{t-1} (q^m - q^i)} \\ &> \prod_{i=0}^{t-1} (1 - q^{i-v}) \geq (1 - q^{t-1-v})^t \geq 1 - \frac{t}{q^{1+v-t}}. \end{aligned}$$

The same analysis holds for $P_2 = P[\text{rank } B_1 = t]$. By the union bound, it follows that the probability of failure satisfies

$$P_f < \frac{2t}{q^{1+v-t}}. \quad (12)$$

Proposition 5: The coding scheme described above can achieve both capacity expressions (8) and (9).

Proof: From (12), we see that achieving either of the limiting capacities amounts to setting a suitable v . To achieve (8), we set $v = t$ and let q grow. The resulting code will have the correct rate, namely, $R = (n - t)(m - t)$ in q -ary units, while the probability of failure will decrease exponentially with the field size in bits.

Alternatively, to achieve (9), we can choose some small $\epsilon > 0$ and set $v = (\tau + \epsilon)n$, where both $\tau = t/n$ and $\lambda = n/m$ are assumed fixed. By letting m grow, we obtain a probability of failure that decreases exponentially with m . The (normalized) gap to capacity of the resulting code will be

$$\begin{aligned} \bar{g} &\triangleq \lim_{m \rightarrow \infty} \overline{C_{\text{AMC}}} - R/(nm) \\ &= (1 - \lambda\tau)(1 - \tau) - (1 - \lambda(\tau + \epsilon))(1 - (\tau + \epsilon)) \\ &= \lambda\epsilon(1 - (\tau + \epsilon)) + \epsilon(1 - \lambda\tau) \\ &< \lambda\epsilon + \epsilon = (1 + \lambda)\epsilon \end{aligned}$$

which can be made as small as we wish. ■

V. THE ADDITIVE-MULTIPLICATIVE MATRIX CHANNEL

Consider a (q, n, m, t) channel with $A \in \mathcal{T}_{n \times n}$, $D \in \mathcal{T}_{n \times t}$ and $Z \in \mathcal{T}_{t \times m}$ uniformly distributed and independent from other variables. Since A is invertible, we can rewrite (2) as

$$Y = AX + DZ = A(X + A^{-1}DZ). \quad (13)$$

Now, since $\mathcal{T}_{n \times n}$ acts transitively on $\mathcal{T}_{n \times t}$, the channel law (13) is equivalent to

$$Y = A(X + W) \quad (14)$$

where $A \in \mathcal{T}_{n \times n}$ and $W \in \mathcal{T}_{n \times m, t}$ are chosen uniformly at random and independently from any other variables. We call (14) *the additive-multiplicative matrix channel (AMMC)*.

A. Capacity

One of the main results of this section is the following theorem, which provides an upper bound on the capacity of the AMMC.

Theorem 6: For $n \leq m/2$, the capacity of the AMMC is upper bounded by

$$C_{\text{AMMC}} \leq (m - n)(n - t) + \log_q 4(1 + n)(1 + t).$$

Proof: Let $S = X + W$. By expanding $I(X, S; Y)$, and using the fact that X , S and Y form a Markov chain, in that order, we have

$$\begin{aligned} I(X; Y) &= I(S; Y) - I(S; Y|X) + \underbrace{I(X; Y|S)}_{=0} \\ &= I(S; Y) - I(W; Y|X) \\ &= I(S; Y) - H(W|X) + H(W|X, Y) \\ &= I(S; Y) - H(W) + H(W|X, Y) \end{aligned} \quad (15)$$

$$\leq C_{\text{MMC}} - \log_q |\mathcal{T}_{n \times m, t}| + H(W|X, Y) \quad (16)$$

where (15) follows since X and W are independent.

We now compute an upper bound on $H(W|X, Y)$. Let $R = \text{rank } Y$ and write $Y = G\bar{Y}$, where $G \in \mathcal{T}_{n \times R}$ and $\bar{Y} \in \mathcal{T}_{R \times m}$. Note that

$$X + W = A^{-1}Y = A^{-1}G\bar{Y} = A^*\bar{Y}$$

where $A^* = A^{-1}G$. Since \bar{Y} is full-rank, it must contain an invertible $R \times R$ submatrix. By reordering columns if necessary, assume that the left $R \times R$ submatrix of \bar{Y} is invertible. Write $\bar{Y} = \begin{bmatrix} \bar{Y}_1 & \bar{Y}_2 \end{bmatrix}$,

$X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ and $W = \begin{bmatrix} W_1 & W_2 \end{bmatrix}$, where \bar{Y}_1 , X_1 and W_1 have R columns, and \bar{Y}_2 , X_2 and W_2 have $m - R$ columns. We have

$$A^* = (X_1 + W_1)\bar{Y}_1^{-1} \quad \text{and} \quad W_2 = A^*\bar{Y}_2 - X_2.$$

It follows that W_2 can be computed if W_1 is known. Thus,

$$\begin{aligned} H(W|X, Y) &= H(W_1|X, Y) \leq H(W_1|R) \leq H(W_1|R = n) \\ &\leq \log_q \sum_{i=0}^t |\mathcal{T}_{n \times n, i}| \leq \log_q (t+1) |\mathcal{T}_{n \times n, t}| \end{aligned} \quad (17)$$

where (17) follows since W_1 may possibly be any $n \times n$ matrix with $\text{rank} \leq t$.

Applying this result in (16), and using (5) and (10), we have

$$\begin{aligned} I(X, Y) &\leq \log_q(n+1) \begin{bmatrix} m \\ n \end{bmatrix} + \log_q(t+1) \frac{|\mathcal{T}_{n \times t}| \begin{bmatrix} n \\ t \end{bmatrix}}{|\mathcal{T}_{n \times t}| \begin{bmatrix} m \\ t \end{bmatrix}} \\ &\leq \log_q(n+1)(t+1) \begin{bmatrix} m-t \\ n-t \end{bmatrix} \\ &\leq (m-n)(n-t) + \log_q 4(1+n)(1+t). \end{aligned} \quad (18)$$

where (18) follows from $\begin{bmatrix} m \\ n \end{bmatrix} \begin{bmatrix} n \\ t \end{bmatrix} = \begin{bmatrix} m \\ t \end{bmatrix} \begin{bmatrix} m-t \\ n-t \end{bmatrix}$, for $t \leq n \leq m$. ■

We now develop a connection with the subspace approach of [9] that will be useful to obtain a lower bound on the capacity. From Section III, we know that, in a multiplicative matrix channel, the receiver can only distinguish between transmitted subspaces. Thus, we can equivalently express

$$C_{\text{AMMC}} = \max_{p_X} I(\mathcal{X}; \mathcal{Y})$$

where \mathcal{X} and \mathcal{Y} denote the row spaces of X and Y , respectively.

Using this interpretation, we can obtain the following lower bound on capacity.

Theorem 7: Assume $n \leq m$. For any $\epsilon \geq 0$, we have

$$C_{\text{AMMC}} \geq (m-n)(n-t-\epsilon t) - \log_q 4 - \frac{2tnm}{q^{1+\epsilon t}}.$$

In order to prove Theorem 7, we need a few lemmas.

Lemma 8: Let $X \in \mathbb{F}_q^{n \times m}$ be a matrix of rank k , and let $W \in \mathbb{F}_q^{n \times m}$ be a random matrix chosen uniformly among all matrices of rank t . If $k+t \leq \min\{n, m\}$, then

$$P[\text{rank}(X+W) < k+t] < \frac{2t}{q^{\min\{n, m\} - k - t + 1}}.$$

Proof: Write $X = X'X''$, where $X' \in \mathbb{F}_q^{n \times k}$ and $X'' \in \mathbb{F}_q^{k \times m}$ are full-rank matrices. We can generate W as $W = W'W''$, where $W' \in \mathcal{T}_{n \times t}$ and $W'' \in \mathcal{T}_{t \times m}$ are chosen uniformly at random and independently from each other. Then we have

$$X + W = X'X'' + W'W'' = \begin{bmatrix} X' & W' \end{bmatrix} \begin{bmatrix} X'' \\ W'' \end{bmatrix}.$$

Note that $\text{rank}(X + W) = k + t$ if and only if the column spaces of X' and W' intersect trivially and the row spaces of X'' and W'' intersect trivially. Let P' and P'' denote the probabilities of these two events, respectively. By a simple counting argument, we have

$$\begin{aligned} P' &= \frac{(q^n - q^k) \cdots (q^n - q^{k+t-1})}{(q^n - 1) \cdots (q^n - q^{t-1})} = \prod_{i=0}^{t-1} \frac{(1 - q^{k-n+i})}{(1 - q^{-n+i})} \\ &> \prod_{i=0}^{t-1} (1 - q^{k-n+i}) \geq (1 - q^{k-n+t-1})^t \geq 1 - tq^{k-n+t-1}. \end{aligned}$$

Similarly, we have $P'' > 1 - tq^{k-m+t-1}$. Thus,

$$\begin{aligned} P[\text{rank}(X + W) < k + t] &< \frac{t}{q^{n-k-t+1}} + \frac{t}{q^{m-k-t+1}} \\ &\leq \frac{2t}{q^{\min\{n,m\}-k-t+1}}. \end{aligned}$$

■

For $\dim \mathcal{X} \leq n \leq m$, let $\mathcal{S}_{\mathcal{X},n}$ denote the set of all n -dimensional subspaces of \mathbb{F}_q^m that contain a subspace $\mathcal{X} \subseteq \mathbb{F}_q^m$.

Lemma 9:

$$|\mathcal{S}_{\mathcal{X},n}| = \begin{bmatrix} m - k \\ n - k \end{bmatrix}_q$$

where $k = \dim \mathcal{X}$.

Proof: By the fourth isomorphism theorem [17], there is a bijection between $\mathcal{S}_{\mathcal{X},n}$ and the set of all $(n - k)$ -dimensional subspaces of the quotient space $\mathbb{F}_q^m / \mathcal{X}$. Since $\dim \mathbb{F}_q^m / \mathcal{X} = m - k$, the result follows. ■

We can now give a proof of Theorem 7.

Proof of Theorem 7: Assume that X is selected from $\mathcal{T}_{n \times m, k}$, where $k = n - (1 + \epsilon)t$ and $\epsilon \geq 0$. Define a random variable Q as

$$Q = \begin{cases} 1 & \text{if } \dim \mathcal{Y} = \text{rank}(X + W) = k + t \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathcal{X} \subseteq \mathcal{Y}$ when $Q = 1$.

By Lemma 9 and (6), we have

$$H(\mathcal{Y}|\mathcal{X}, Q = 1) \leq \log_q |\mathcal{S}_{\mathcal{X}, n'}| \leq (m - n')t + \log_q 4$$

where $n' = k + t$. Choosing X uniformly from $\mathcal{T}_{n \times m, k}$, we can also make \mathcal{Y} uniform within a given dimension; in particular,

$$H(\mathcal{Y}|Q = 1) = \log_q \left[\begin{matrix} m \\ n' \end{matrix} \right]_q \geq (m - n')n'.$$

It follows that

$$\begin{aligned} I(\mathcal{X}; \mathcal{Y}|Q = 1) &= H(\mathcal{Y}|Q = 1) - H(\mathcal{Y}|\mathcal{X}, Q = 1) \\ &\geq (m - n')(n' - t) - \log_q 4 \\ &\geq (m - n)(n - t - \epsilon t) - \log_q 4. \end{aligned}$$

Now, using Lemma 8, we obtain

$$\begin{aligned} I(\mathcal{X}; \mathcal{Y}) &= I(\mathcal{X}; \mathcal{Y}, Q) = I(\mathcal{X}; Q) + I(\mathcal{X}; \mathcal{Y}|Q) \\ &\geq I(\mathcal{X}; \mathcal{Y}|Q) \\ &\geq P[Q = 1]I(\mathcal{X}; \mathcal{Y}|Q = 1) \\ &\geq I(\mathcal{X}; \mathcal{Y}|Q = 1) - P[Q = 0]nm \\ &\geq (m - n)(n - t - \epsilon t) - \log_q 4 - \frac{2tnm}{q^{\epsilon t + 1}}. \end{aligned}$$

■

Note that, differently from the results of previous sections, Theorems 6 and 7 provide only upper and lower bounds on the channel capacity. Nevertheless, it is still possible to compute exact expressions for the capacity of the AMMC in certain limiting cases.

Corollary 10: For $0 < \lambda = n/m \leq 1/2$ and $\tau = t/n$, we have

$$\lim_{q \rightarrow \infty} C_{\text{AMMC}} = (m - n)(n - t) \tag{19}$$

$$\lim_{\substack{m \rightarrow \infty \\ n = \lambda m \\ t = \tau n}} \overline{C_{\text{AMMC}}} = (1 - \lambda)(1 - \tau). \tag{20}$$

Proof: The fact that the values in (19) and (20) are upper bounds follows immediately from Theorem 6. The fact that (19) is a lower bound follows immediately from Theorem 7 by setting $\epsilon = 0$. To obtain (20) from Theorem 7, it suffices to choose ϵ such that $1/\epsilon$ grows sublinearly with m , e.g., $\epsilon = 1/\sqrt{m}$. ■

Once again, note that (19) agrees with (20) if we consider the normalized capacity.

Differently from the MMC and the AMC, successful decoding in the AMMC does not (necessarily) allow recovery of all sources of channel uncertainty—in this case, the matrices A and W . In general, for every observable (X, Y) pair, there are many valid A and W such that $Y = A(X + W)$. Such coupling between A and W is reflected in extra term $H(W|X, Y)$ in (15), which provides an additional rate of roughly $(2n - t)t$ as compared to the straightforward lower bound $C_{\text{AMMC}} \geq C_{\text{MMC}} - \log_q |\mathcal{T}_{n \times m, t}| \approx (m - n)n - (n + m - t)t$.

Remark: In [11], the problem of communicating over the AMMC was addressed assuming a specific form of transmission matrices that contained an $n \times n$ identity header. Note that, if we consider such a header as being part of the channel (i.e., beyond the control of the code designer) then, with high probability, the resulting channel becomes equivalent to the AMC (see [11] for details). However, as a coding strategy for the AMMC, using such an $n \times n$ identity header results in a suboptimal input distribution, as the mutual information achieved is strictly smaller than the capacity. Indeed, the capacity-achieving distribution used in Theorem 7 and Corollary 10 corresponds to transmission matrices of rank $n - (1 + \epsilon)t$. This result shows that, for the AMMC, using headers is neither fundamental nor asymptotically optimal.

B. A Coding Scheme

We now propose an efficient coding scheme that can asymptotically achieve (19) and (20). The scheme is based on a combination of channel sounding and error trapping strategies.

For a data matrix $U \in \mathbb{F}_q^{(n-v) \times (m-n)}$, where $v \geq t$, let the corresponding codeword be

$$X = \begin{bmatrix} 0 \\ \bar{X} \end{bmatrix} = \begin{bmatrix} 0_{v \times v} & 0_{v \times (n-v)} & 0_{v \times (m-n)} \\ 0_{(n-v) \times v} & I_{(n-v) \times (n-v)} & U \end{bmatrix}.$$

Note that the all-zero matrices provide the error traps, while the identity matrix corresponds to the pilot symbols. Clearly, the rate of this scheme is $R = (n - v)(m - n)$.

Write the noise matrix W as

$$W = BZ = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \begin{bmatrix} Z_1 & Z_2 & Z_3 \end{bmatrix}$$

where $B_1 \in \mathbb{F}_q^{v \times t}$, $B_2 \in \mathbb{F}_q^{(n-v) \times t}$, $Z_1 \in \mathbb{F}_q^{t \times v}$, $Z_2 \in \mathbb{F}_q^{t \times (n-v)}$ and $Z_3 \in \mathbb{F}_q^{t \times (m-n)}$. The auxiliary matrix S is then given by

$$S = X + W = \begin{bmatrix} B_1 Z_1 & B_1 Z_2 & B_1 Z_3 \\ B_2 Z_1 & I + B_2 Z_2 & U + B_2 Z_3 \end{bmatrix}.$$

Similarly as in Section IV, we define that the error trapping is successful if $\text{rank } B_1 Z_1 = t$. Assume that this is the case. From Section IV, there exists some matrix $T \in \mathcal{T}_{n \times n}$ such that

$$TS = \begin{bmatrix} B_1 Z_1 & B_1 Z_2 & B_1 Z_3 \\ 0 & I & U \end{bmatrix} = \begin{bmatrix} B_1 & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} Z_1 & Z_2 & Z_3 \\ 0 & I & U \end{bmatrix}.$$

Note further that

$$\text{RRE} \left(\begin{bmatrix} Z_1 & Z_2 & Z_3 \\ 0 & I & U \end{bmatrix} \right) = \begin{bmatrix} \tilde{Z}_1 & 0 & \tilde{Z}_3 \\ 0 & I & U \end{bmatrix}$$

for some $\tilde{Z}_1 \in \mathbb{F}_q^{t \times v}$ in RRE form and some $\tilde{Z}_3 \in \mathbb{F}_q^{t \times (m-n)}$. It follows that

$$\begin{aligned} \text{RRE}(S) &= \text{RRE} \left(\begin{bmatrix} B_1 & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} Z_1 & Z_2 & Z_3 \\ 0 & I & U \end{bmatrix} \right) \\ &= \begin{bmatrix} \tilde{Z}_1 & 0 & \tilde{Z}_3 \\ 0 & I & U \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

where the bottom $v - t$ rows are all-zeros.

Since A is invertible, we have $\text{RRE}(Y) = \text{RRE}(S)$, from which U can be readily obtained. Thus, decoding amounts to performing Gauss-Jordan elimination on Y . It follows that the complexity of the scheme is $O(n^2 m)$ operations in \mathbb{F}_q .

The probability that the error trapping is not successful, i.e., $\text{rank } B_1 Z_1 < t$, was computed in Section IV. Let \hat{A} correspond to the first n columns of Y . Note that $\text{rank } B_1 Z_1 = t$ if and only if $\text{rank } \hat{A} = n - v + t$. Thus, when the error trapping is not successful, the receiver can easily detect this event by looking at $\text{RRE}(Y)$ and then declare a decoding failure. It follows that the scheme has zero error probability and probability of failure given by (12).

Theorem 11: The proposed coding scheme can asymptotically achieve (19) and (20).

Proof: Using (12) and the same argument as in the proof of Proposition 5, we can set a suitable v in order to achieve arbitrarily low gap to capacity while maintaining an arbitrary low probability of failure, for both cases where $q \rightarrow \infty$ or $m \rightarrow \infty$. ■

VI. EXTENSIONS

In this section, we discuss possible extensions of the results and models presented in the previous sections.

A. Dependent Transfer Matrices

As discussed in Section V, the AMMC is equivalent to a channel of the form (2) where $A \in \mathcal{T}_{n \times n}$ and $D \in \mathcal{T}_{n \times t}$ are chosen uniformly at random and independently from each other. Suppose now that the channel is the same, except for the fact that A and D are not independent. It should be clear that the capacity of the channel cannot be smaller than that of the AMMC. For instance, one can always convert this channel into an AMMC by employing randomization at the source. (This is, in fact, a natural procedure in any random network coding system.) Let $X = TX'$, where $T \in \mathcal{T}_{n \times n}$ is chosen uniformly at random and independent from any other variables. Then $A' = AT$ is uniform on $\mathcal{T}_{n \times n}$ and independent from D . Thus, the channel given by $Y = A'X' + DZ$ is an AMMC.

Note that our coding scheme does not rely on any particular statistics of A given X and W (except the assumption that A is invertible) and therefore works unchanged in this case.

B. Transfer Matrix Invertible but Nonuniform

The model for the AMMC assumes that the transfer matrix $A \in \mathcal{T}_{n \times n}$ is chosen uniformly at random. In a realistic network coding system, the transfer matrix may be a function of both the network code and the network topology, and therefore may not have a uniform distribution. Consider the case where A is chosen according to an arbitrary probability distribution on $\mathcal{T}_{n \times n}$. It should be clear that the capacity can only increase as compared with the AMMC, since less “randomness” is introduced in the channel. The best possible situation is to have a constant A , in which case the channel becomes exactly an AMC.

Again, note that our coding scheme for the AMMC is still applicable in this case.

C. Nonuniform Packet Errors

When expressed in the form (2), the models for both the AMC and the AMMC assume that the matrix Z is uniformly distributed on $\mathcal{T}_{n \times t}$. In particular, each error packet is uniformly distributed on $\mathbb{F}_q^{1 \times m} \setminus \{0\}$. In a realistic situation, however, it may be the case that error packets of low weight are more likely to occur. Consider a model identical to the AMC or the AMMC except for the fact that the matrix Z is chosen according to an arbitrary probability distribution on $\mathcal{T}_{t \times m}$. Once again, it should be clear that the capacity can only increase. Note that the exact capacity in Proposition 4 and the upper bound of Theorem 6 can be easily modified to account for this case (by replacing $\log_q |\mathcal{T}_{n \times m, t}|$ with the entropy of W).

Although our coding scheme in principle does not hold in this more general case, we can easily convert the channel into an AMC or AMMC by applying a random transformation at the source (and its inverse

at the destination). Let $X = X'T$, where $T \in \mathcal{T}_{m \times m}$ is chosen uniformly at random and independent from any other variables. Then

$$Y' = YT^{-1} = (AX + DZ)T^{-1} = AX' + DZ'$$

where $Z' = ZT^{-1}$. Since $\mathcal{T}_{m \times m}$ acts (by right multiplication) transitively on $\mathcal{T}_{t \times m}$, we have that Z' is uniform on $\mathcal{T}_{t \times m}$. Thus, we obtain precisely an AMMC (or AMC) and the assumptions of our coding scheme hold.

Note, however, that, depending on the error model, the capacity may be much larger than what can be achieved by the scheme described above. For instance, if the rows of Z are constrained to have weight at most s (otherwise chosen, say, uniformly at random), then the capacity would increase by approximately $(m - s - \log_q \binom{m}{s})t$, which might be a substantial amount if s is small.

D. Error Matrix with Variable Rank ($\leq t$)

The model we considered for the AMC and the AMMC assumes an error matrix W whose rank is known and equal to t . It is useful to consider the case where rank W is allowed to vary, while still bounded by t . More precisely, we assume that W is chosen uniformly at random from $\mathcal{T}_{n \times m, R}$, where $R \in \{0, \dots, t\}$ is a random variable with probability distribution $P[R = r] = p_r$.

Since

$$\begin{aligned} H(W) &= H(W, R) = H(R) + H(W|R) \\ &= H(R) + \sum_r p_r H(W|R = r) \\ &= H(R) + \sum_r p_r \log_q |\mathcal{T}_{n \times m, r}| \\ &\leq H(R) + \log_q |\mathcal{T}_{n \times m, t}|, \end{aligned}$$

we conclude that the capacities of the AMC and the AMMC may be reduced by at most $H(R) \leq \log_q(t+1)$. This loss is asymptotically negligible for large q and/or large m , so the expressions (8), (9), (19) and (20) remain unchanged.

The steps for decoding and computing the probability of error trapping failure also remain the same, provided we replace t by R . The only difference is that now decoding errors may occur. More precisely, suppose that $\text{rank } B_1 Z_1 = t' < t$. A necessary condition for success is that $\text{rank } B_1 Z = \text{rank } B Z_1 = t'$. If this condition is not satisfied, then a decoding failure is declared. However, if the condition is true, then the decoder cannot determine whether $t' = R < t$ (an error trapping success) or $t' < R \leq t$ (an

error trapping failure), and must proceed assuming the former case. If the latter case turns out to be true, we would have an undetected error. Thus, for this model, the expression (12) gives a bound on the probability that decoding is not successful, i.e., that either an error or a failure occurs.

E. Infinite-Packet-Length Channel and Infinite-Batch-Size Channel

We now extend our results to the infinite-packet-length AMC and AMMC and the infinite-batch-size AMC. (Note that, as pointed out in Section III, there is little justification to consider an infinite-batch-size AMMC.) From the proof of Proposition 4 and the proof of Corollary 10, it is straightforward to see that

$$\begin{aligned}\lim_{m \rightarrow \infty} \overline{C_{\text{AMMC}}} &= \lim_{m \rightarrow \infty} \overline{C_{\text{AMC}}} = (n - t)/n \\ \lim_{n \rightarrow \infty} \overline{C_{\text{AMC}}} &= (m - t)/m.\end{aligned}$$

It is *not* straightforward, however, to obtain capacity-achieving schemes for these channels. The schemes described in Sections IV and V for the infinite-rank AMC and AMMC, respectively, use an error trap whose size (in terms of columns *and* rows) grows proportionally with m (or n). While this is necessary for achieving vanishingly small error probability, it also implies that these schemes are not suitable for the infinite-packet-length channel (where $m \rightarrow \infty$ but not n) or the infinite-batch-size channel (where $n \rightarrow \infty$ but not m).

In these situations, the proposed schemes can be adapted by replacing the data matrix and part of the error trap with a *maximum-rank-distance* (MRD) code [18]. Consider first an infinite-packet-length AMC. Let the transmitted matrix be given by

$$X = \begin{bmatrix} 0_{n \times v} & x \end{bmatrix} \quad (21)$$

where $x \in \mathbb{F}_q^{n \times (m-v)}$ is a codeword of a matrix code \mathcal{C} . If (column) error trapping is successful then, under the terminology of [10], the decoding problem for \mathcal{C} amounts to the correction of t *erasures*. It is known that, for $m - v \geq n$, an MRD code $\mathcal{C} \subseteq \mathbb{F}_q^{n \times (m-v)}$ with rate $(n - t)/n$ can correct exactly t erasures (with zero probability of error) [10]. Thus, decoding fails if and only if column trapping fails.

Similarly, for an infinite-batch-size AMC, let the transmitted matrix be given by

$$X = \begin{bmatrix} 0_{v \times m} \\ x \end{bmatrix}$$

where $x \in \mathbb{F}_q^{(n-v) \times m}$ is a codeword of a matrix code \mathcal{C} . If (row) error trapping is successful then, under the terminology of [10], the decoding problem for \mathcal{C} amounts to the correction of t *deviations*. It is known

that, for $n - v \geq m$, an MRD code $\mathcal{C} \subseteq \mathbb{F}_q^{(n-v) \times m}$ with rate $(m - t)/m$ can correct exactly t deviations (with zero probability of error) [10]. Thus, decoding fails if and only if row trapping fails.

Finally, for the infinite-packet-length AMMC, it is sufficient to prepend to (21) an identity matrix, i.e.,

$$X = \begin{bmatrix} I_{n \times n} & 0_{n \times v} & x \end{bmatrix}.$$

The same reasoning as for the infinite-packet-length AMC applies here, and the decoder in [10] is also applicable in this case.

For more details on the decoding of an MRD code combined with an error trap, we refer the reader to [19]. The decoding complexity is in $O(tn^2m)$ and $O(tm^2n)$ (whichever is smaller) [10].

In all cases, the schemes have probability of error upper bounded by t/q^{1+v-t} and therefore are capacity-achieving.

VII. CONCLUSIONS

We have considered the problem of reliable communication over certain additive matrix channels inspired by network coding. These channels provide a reasonable model for both coherent and random network coding systems subject to random packet errors. In particular, for an additive-multiplicative matrix channel, we have obtained upper and lower bounds on capacity for any channel parameters and asymptotic capacity expressions in the limit of large field size and/or large matrix size; roughly speaking, we need to use t redundant packets in order to be able to correct up to t injected error packets. We have also presented a simple coding scheme that achieves capacity in these limiting cases while requiring a significantly low decoding complexity; in fact, decoding amounts simply to performing Gauss-Jordan elimination, which is already the standard decoding procedure for random network coding. Compared to previous work on correction of adversarial errors (where approximately $2t$ redundant packets are required), the results of this paper show an improvement of t redundant packets that can be used to transport data, if errors occur according to a probabilistic model.

Several questions remain open and may serve as an interesting avenue for future research:

- Our results for the AMMC assume that the transfer matrix A is always nonsingular. It may be useful to consider a model where $\text{rank } A$ is a random variable. Note that, in this case, one cannot expect to achieve reliable (and efficient) communication with a one-shot code, as the channel realization would be unknown at the transmitter. Thus, in order to achieve capacity under such a model (even with arbitrarily large q or m), it is strictly necessary to consider multi-shot codes.

- As pointed out in Section VI-C, our proposed coding scheme may not be even close to optimal when packet errors occur according to a nonuniform probability model. Especially in the case of low-weight errors, it is an important question how to approach capacity with a low-complexity coding scheme. It might also be interesting to know whether one-shot codes are still useful in this case.
- Another important assumption of this paper is the bounded number of $t < n$ packet errors. What if t is unbounded (although with a low number of errors being more likely than a high number)? While the capacity of such a channel may not be too hard to approximate (given the results of this paper), finding a low-complexity coding scheme seems a very challenging problem.

ACKNOWLEDGEMENTS

We would like to thank the associate editor and the anonymous reviewers for their helpful comments.

REFERENCES

- [1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
- [2] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.
- [3] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.
- [4] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.
- [5] D. Silva and F. R. Kschischang, "On metrics for error correction in network coding," *IEEE Trans. Inf. Theory*, 2008, to be published. [Online]. Available: <http://arxiv.org/abs/0805.3824>
- [6] —, "Adversarial error correction for network coding: Models and metrics," in *Proc. 46th Annual Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, Sep. 23–26, 2008, pp. 1246–1253.
- [7] N. Cai and R. W. Yeung, "Network error correction, part II: Lower bounds," *Commun. Inform. Syst.*, vol. 6, no. 1, pp. 37–54, 2006.
- [8] R. W. Yeung and N. Cai, "Network error correction, part I: Basic concepts and upper bounds," *Commun. Inform. Syst.*, vol. 6, no. 1, pp. 19–36, 2006.
- [9] R. Kötter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3579–3591, Aug. 2008.
- [10] D. Silva, F. R. Kschischang, and R. Kötter, "A rank-metric approach to error control in random network coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3951–3967, 2008.
- [11] A. Montanari and R. Urbanke, "Iterative coding for network coding," 2007, submitted for publication. [Online]. Available: <http://arxiv.org/abs/0711.3935>
- [12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & sons, 1991.

- [13] M. Siavoshani, C. Fragouli, and S. Diggavi, "Noncoherent multisource network coding," in *Proc. IEEE Int. Symp. Information Theory*, Toronto, Canada, Jul. 6–11, 2008, pp. 817–821.
- [14] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, M. Médard, and M. Effros, "Resilient network coding in the presence of Byzantine adversaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2596–2603, Jun. 2008.
- [15] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. Allerton Conf. on Comm., Control, and Computing*, Monticello, IL, Oct. 2003, pp. 40–49.
- [16] R. Lidl and H. Niederreiter, *Finite Fields*. Reading, MA: Addison-Wesley, 1983.
- [17] D. S. Dummit and R. M. Foote, *Abstract Algebra*, 3rd ed. John Wiley & Sons, 2004.
- [18] E. M. Gabidulin, "Theory of codes with maximum rank distance," *Probl. Inform. Transm.*, vol. 21, no. 1, pp. 1–12, 1985.
- [19] D. Silva and F. R. Kschischang, "A key-based error control scheme for network coding," in *Proc. 11th Canadian Workshop Inform. Theory*, Ottawa, Canada, May 13–15, 2009, pp. 5–8.

Danilo Silva (S'06) received the B.Sc. degree from the Federal University of Pernambuco, Recife, Brazil, in 2002, the M.Sc. degree from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil, in 2005, and the Ph.D. degree from the University of Toronto, Toronto, Canada, in 2009, all in electrical engineering.

He is currently a Postdoctoral Fellow at the University of Toronto. His research interests include channel coding, information theory, and network coding.

Frank R. Kschischang (S'83–M'91–SM'00–F'06) received the B.A.Sc. degree (with honors) from the University of British Columbia, Vancouver, BC, Canada, in 1985 and the M.A.Sc. and Ph.D. degrees from the University of Toronto, Toronto, ON, Canada, in 1988 and 1991, respectively, all in electrical engineering.

He is a Professor of Electrical and Computer Engineering and Canada Research Chair in Communication Algorithms at the University of Toronto, where he has been a faculty member since 1991. During 1997–1998, he was a Visiting Scientist at the Massachusetts Institute of Technology, Cambridge, and in 2005 he was a Visiting Professor at the ETH, Zürich, Switzerland. His research interests are focused on the area of channel coding techniques.

Prof. Kschischang was the recipient of the Ontario Premier's Research Excellence Award. From 1997 to 2000, he served as an Associate Editor for Coding Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY. He also served as Technical Program Co-Chair for the 2004 IEEE International Symposium on Information Theory (ISIT), Chicago, IL, and as General Co-Chair for ISIT 2008, Toronto.

Ralf Kötter (S'92-M'95-SM'06-F'09) received a Diploma in Electrical Engineering from the Technical University Darmstadt, Germany in 1990 and a Ph.D. degree from the Department of Electrical Engineering at Linköping University, Sweden.

From 1996 to 1997, he was a Visiting Scientist at the IBM Almaden Research Laboratory in San Jose, CA. He was a Visiting Assistant Professor at the University of Illinois at Urbana-Champaign and a Visiting Scientist at CNRS in Sophia-Antipolis, France, from 1997 to 1998. In the years 1999-2006 he was member of the faculty of the University of Illinois at Urbana-Champaign, where his research interests included coding and information theory and their application to communication systems. In 2006 he joined the faculty of the Technische Universität München, Munich, Germany, as the Head of the Institute for Communications Engineering.

He served as an Associate Editor for both the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON INFORMATION THEORY. He received an IBM Invention Achievement Award in 1997, an NSF CAREER Award in 2000, an IBM Partnership Award in 2001, and a 2006 XEROX award for faculty research. He is co-recipient of the 2004 Information Theory Society Best Paper Award, of the 2004 IEEE Signal Processing Magazine Best Paper Award, and of the 2009 Joint Communications Society and Information Theory Society Best Paper Award. He received the Vodafone Innovationspreis in 2008.

Ralf Kötter died in February, 2009.